Distinct haplotype structure at the innate immune receptor Toll-like receptor 2 (TLR2) across bank vole populations and lineages in Europe

Morger, Jennifer ; Raberg, L ; Hille, S M ; Helsen, S ; Štefka, J ; Al-Sabi, M M ; Kapel, C M O ; Mappes, T ; Essbauer, S ; Ulrich, R G ; Bartolommei, P ; Mortelliti, A ; Balčiauskas, L ; van den Brink, N W ; Rémy, A ; Bajer, A ; Cheprakov, M ; Korva, M ; García-Pérez, A L ; Biek, R ; Withenshaw, S ; Tschirren, Barbara

Abstract: Parasite-mediated selection may contribute to the maintenance of genetic variation at host immune genes over long time scales. To date, the best evidence for the long-term maintenance of immunogenetic variation in natural populations comes from studies on the major histocompatibility complex (MHC) genes, whereas evidence for such processes from other immune genes remains scarce. In the present study, we show that, despite pronounced population differentiation and the occurrence of numerous private alleles within populations, the innate immune gene Toll-like receptor 2 (TLR2) displays a distinct haplotype structure in 21 bank vole (Myodes glareolus) populations across Europe. Haplotypes from all populations grouped in four clearly differentiated clusters, with the three main clusters co-occurring in at least three previously described mitochondrial lineages. This pattern indicates that the distinct TLR2 haplotype structure may precede the split of the mitochondrial lineages 0.19–0.56 Mya and suggests that haplotype clusters at this innate immune receptor are maintained over prolonged time in wild bank vole populations.

DOI: https://doi.org/10.1111/bij.12593

Distinct haplotype structure at the innate immune receptor Toll-like receptor 2 (*TLR2*) across bank vole populations and lineages in Europe

Jennifer Morger[1], Lars Råberg[2], Sabine M. Hille[3], Sanne Helsen[4], Jan Štefka[5], Mohammad M. Al-Sabi[6], Christian M. O. Kapel[6], Tapio Mappes[7], Sandra Essbauer[8], Rainer G. Ulrich[9], Paola Bartolommei[10], Alessio Mortelliti[11,12], Linas Balčiauskas[13], Nico W. van den Brink[14,15], Alice Rémy[16], Anna Bajer[17], Mihail Cheprakov[18], Misa Korva[19], Ana L. García-Pérez[20], Roman Biek[21], Susan Withenshaw[22] & Barbara Tschirren[1]

[1] Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland

[2] Department of Biology, Lund University, Sölvegatan 35, 223 62 Lund, Sweden

[3] Institute of Wildlife Biology and Game Management, University of Natural Resources and Life Sciences, Gregor Mendel-Strasse 33,1180 Vienna, Austria

[4] Evolutionary Ecology Group, Department of Biology, University of Antwerp, Groenenborgerlaan 171, 2020 Antwerp, Belgium

[5] Biology Centre ASCR, Institute of Parasitology and University of South Bohemia, Faculty of Science, Branišovská 31, 37005, České Budějovice, Czech Republic

[6] Department of Plant and Environmental Sciences, Faculty of Science, University of Copenhagen, Thorvaldsensvej 40, 1871 Frederiksberg C, Denmark

[7] Department of Biological and Environmental Science, University of Jyväskylä, PO Box 35, 40014 Jyväskylä, Finland

[8] Bundeswehr Institute of Microbiology, Department Virology and Rickettsiology, Neuherbergstr. 11, 80937 Munich, Germany

[9] Friedrich-Loeffler-Institut, Federal Research Institute for Animal Health, Institute for Novel and Emerging Infectious Diseases, Südufer 10, 17493 Greifswald - Insel Riems, Germany

[10] Fondazione Ethoikos, Convento dell'Osservanza, 53030 Radicondoli, Siena, Italy

[11] Department of Biology and Biotechnology "Charles Darwin", University of Rome "La Sapienza", Viale dell'Università 32, 00185, Rome, Italy

[12] Fenner School of Environment and Society, Australian Research Council Centre for Environmental Decisions, National Environmental Research Program, The Australian National University, Canberra, ACT 0200, Australia

[13] Nature Research Centre, Akademijos 2, 08412 Vilnius, Lithuania

[14] Alterra, Wageningen UR. PO-Box 47, 6700 AA Wageningen, The Netherlands

[15] Wageningen University Wageningen UR. Sub-department of Toxicology. PO-Box 8000, 6700 EA, Wageningen, The Netherlands

[16] Faculty of Applied Ecology and Agricultural Sciences, Hedmark University College, Anne Evenstadsvei 80, 2480 Koppang, Norway

[17] Department of Parasitology, Institute of Zoology, Faculty of Biology, University of Warsaw,1 Miecznikowa Street, 02-096, Warsaw, Poland

[18] Institute of Plant and Animal Ecology, Ural Branch, Russian Academy of Sciences, Str 8 Marta 202, Yekaterinburg, 620144, Russia

[19] Institute of Microbiology and Immunology, Faculty of Medicine, Zaloška 4, 1000 Ljubljana, Slovenia

[20] NEIKER – Instituto Vasco de Investigación y Desarrollo Agrario, Department of Animal Health, Berreaga 1, 48160 Derio, Bizkaia, Spain

[21] Boyd Orr Centre for Population and Ecosystem Health, Institute of Biodiversity, Animal Health and Comparative Medicine, College of Medical Veterinary and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom

[22] Department of Evolution, Ecology & Behaviour, Institute of Integrative Biology, University of Liverpool, Crown Street, Liverpool, L69 7ZB, United Kingdom

Correspondence: Barbara Tschirren, Institute of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland; email: barbara.tschirren@ieu.uzh.ch, phone: +41 44 635 47 77, Fax: +41 44 635 47 80

Running title: *TLR2* variation in European bank voles

Abstract

Parasite-mediated selection may contribute to the maintenance of genetic variation at host immune genes over long time scales. To date, the best evidence for the long-term maintenance of immunogenetic variation in natural populations comes from studies on the major histocompatibility complex (*MHC*), whereas evidence for such processes from other immune genes remains scarce.

Here we show that despite pronounced population differentiation and the occurrence of numerous private alleles within populations, the innate immune gene Toll-like receptor 2 (*TLR2*) displays a distinct haplotype structure in 21 bank vole (*Myodes glareolus*) populations across Europe. Haplotypes from all populations grouped in four clearly differentiated clusters, with the three main clusters co-occurring in at least three previously described mitochondrial lineages. This pattern indicates that the distinct *TLR2* haplotype structure may precede the split of the mitochondrial lineages 0.19 – 0.56 Mya and suggests that haplotype clusters at this innate immune receptor are maintained over prolonged time in wild bank vole populations.

Keywords: balancing selection – diversity – immune gene evolution – immunogenetics – maintenance of genetic variation – *Myodes glareolus* – parasite resistance – phylogeography – rodents – Toll-like receptors

Introduction

Understanding how genetic variation is maintained in natural populations is a fundamental quest in evolutionary biology. Immune genes frequently display particularly high diversity, presumably as a result of parasite-mediated selection (Haldane, 1949). There are different, non-mutually exclusive processes by which parasites (i.e. parasitic and pathogenic organisms and agents) may contribute to the maintenance of genetic variation at host defence genes. Firstly, parasites may specialise in the most common host genotype, leading to negative frequency-dependent selection (Takahata & Nei, 1990; Woolhouse *et al.*, 2002). Secondly, heterozygous individuals may have a selective advantage over homozygous individuals (i.e. overdominance), for example because of enhanced parasite recognition ability or reduced immunopathology (Doherty & Zinkernagel, 1975; Apanius *et al.*, 1997; Khor *et al.*, 2007). Finally, variation in the composition of the parasite community in time and / or space may lead to fluctuating selection, and hence contribute to the maintenance of genetic variation in host populations (Hedrick, 2002).

To date, the best evidence for the maintenance of immunogenetic variation over long time scales in natural populations comes from studies on the major histocompatibility complex (*MHC*) genes (Hedrick, 1998; Bernatchez & Landry, 2003; Piertney & Oliver, 2006). In humans, a number of other immune loci with signatures of long-term balancing selection have been identified (Bubb *et al.*, 2006, Andrés *et al.* 2009). In other species, however,

114 evidence for such patterns outside the MHC complex remains scarce (but see
115 e.g. Ferguson *et al.*, 2012).
116 Toll-like receptors (TLRs) are pattern-recognition receptors (PRRs) that
117 play a key role in the recognition of intra- and extracellular pathogens. They
118 recognise essential structures of pathogens (Pathogen-Associated Molecular
119 Patterns; PAMPs) and initiate innate and adaptive immune responses
120 (Medzhitov, 2001; Akira, Uematsu & Takeuchi, 2006). Toll-like receptor 2
121 (*TLR2*), for example, plays an important role in the recognition of lipopeptides,
122 crucial components of the cell membranes of Gram-positive bacteria (Jin *et*
123 *al.*, 2007). Genetic polymorphisms at *TLRs* have been associated with
124 variation in infectious disease susceptibility in humans and laboratory animals
125 (Texereau *et al.*, 2005; Netea, Wijmenga & O'Neill, 2012), and there is
126 accumulating evidence that they play an important role in mediating host-
127 parasite interactions and disease susceptibility in wildlife (Jackson *et al.*,
128 2009; Turner *et al.*, 2011; Tschirren *et al.*, 2013; Morger *et al*,. 2014;
129 Fornůsková *et al.*, 2014).
130 It has previously been found that in a Swedish bank vole *Myodes*
131 *glareolus* (Schreber, 1780; Rodentia, Cricetidae) population *TLR2* displays a
132 high diversity, with haplotypes grouping in three distinct clusters (Tschirren *et*
133 *al.*, 2013). Furthermore, voles carrying haplotypes of one particular cluster (c_2)
134 had a reduced *Borrelia afzelii* infection rate, and an analysis of haplotype
135 frequencies and tree topology strongly suggested that these 'protective'
136 alleles had evolved under positive selection (Tschirren *et al.*, 2013). One
137 interpretation of these patterns is that the high diversity at *TLR2* is the result

of parasite-mediated selection. However, the distinct *TLR2* haplotype structure could also be a result of neutral processes, such as drift in isolated populations followed by population admixture. Indeed, eight well-defined mitochondrial bank vole lineages have been described in Europe (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006; Deffontaine *et al.*, 2009), and there is evidence that in Southern Fennoscandia re-colonisation after the last glaciation (around 11,000 – 10,000 years BP; Jaarola, Tegelström & Fredga, 1999) occurred from separate glacial refugia, creating contact zones between bank vole lineages that have evolved genetic differences while geographically isolated (Hewitt 1999; Jaarola, Tegelström & Fredga, 1999).

Here we sequenced *TLR2* from 21 bank vole populations across Europe, belonging to five previously described mitochondrial lineages, to test if we find evidence that the different *TLR2* clusters are associated with different bank vole lineages. If, on the other hand, the co-occurrence of the distinct *TLR2* clusters is a general feature of this innate immune gene, and observed in several lineages and populations on a large geographical scale, this would provide indirect evidence for the long-term maintenance of *TLR2* diversity.

Materials and Methods

Study species and sampling sites

The bank vole is a common rodent in Europe. Its distribution ranges from Fennoscandia and the British Isles to northern Spain, Italy, and Russia (Corbet & Harris, 1991). For our study, we caught 385 bank voles at 21

locations in 18 countries for tissue or hair collection (mean per location: 18.3 animals; range: 10 - 21; Table 1, S1). Trapping and tissue or hair sampling were performed in accordance with the regulations of the respective countries and performed under licenses provided by the local animal welfare committees (Table S1). Distances between locations ranged from 69 to 4454 km.

Toll-like receptor 2 (TLR2) and cytochrome b (cytb) genotyping

Total genomic DNA was extracted from the samples using a standard extraction method (Laird *et al.*, 1991). We sequenced a 1149 bp fragment of *TLR*2 for all individuals (*N* = 385, Table 1) following previously published protocols (Tschirren *et al.*, 2012). The amplified fragment contains the functionally relevant sites involved in pathogen-recognition and TLR heterodimerization (Jin *et al.*, 2007; Figure S2), and we previously demonstrated molecular signatures of positive selection during the evolutionary history of rodents within this gene region (Tschirren, Råberg & Westerdahl, 2011). *TLR2* amplifications were performed in a total volume of 10 µl containing 0.2 µl JumpStart Taq DNA Polmyerase (Sigma-Aldrich), 300 nM of the primers MglTLR2F: CATCCATCACCTGACCCTTC and MglTLR2R: CCAGTAGGAATCCTGCTCG and 25 ng of genomic DNA.

To assign the different populations to the previously described mitochondrial bank vole lineages (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006), we sequenced a 1101 bp fragment of the mitochondrial cytochrome b *(cytb)* gene for three to five individuals per population following Kotlík *et al.* (2006).

186 Amplifications were performed in a total volume of 10 µl containing 0.2 µl
187 JumpStart Taq DNA Polymerase (Sigma-Aldrich), 300 nM of each primer and
188 25 ng of genomic DNA. The PCR protocol for *TLR2* amplification consisted of
189 an initial denaturation step at 94°C for 1 min, followed by 35 cycles of
190 denaturation at 94°C for 30 sec, annealing at 57°C for 30 sec, and extension
191 at 72°C for 90 sec, with a final elongation step at 72°C for 10 min. For *cytb* we
192 used 40 cycles and an annealing temperature of 52°C. PCR products were
193 purified and sequenced in both directions on an ABI Prism 3730 capillary
194 sequencer (Applied Biosystems) using Big Dye terminator version 3.1
195 chemistry (Applied Biosystems). All *TLR2* and *cytb* raw sequences were
196 trimmed, processed and aligned in Geneious 5.6.5 (Drummond *et al.*, 2009).
197 Putative polymorphisms identified by the program were manually verified.
198 *TLR2* haplotypes were reconstructed for each population separately in the
199 program PHASE 2.1 (Stephens, Smith & Donnelly, 2001) using the default
200 settings of a thinning interval of one, 100 burn-in iterations and 100 main
201 iterations, and allowing for recombination (see Table S3 for additional
202 information). If a haplotype was observed only once in the dataset, we
203 repeated the PCR and sequencing reactions to confirm the sequence and
204 avoid overestimation of diversity due to amplification or sequencing errors.
205 *TLR2* and *cytb* sequences were submitted to GenBank (*TLR2* accession
206 numbers: Table S4; *cytb* accession numbers: KJ612463 - KJ612512).

207

208 *Statistical analyses*

209 We quantified genetic diversity at *TLR2* in the 21 bank vole populations by

calculating the number of haplotypes, the number of synonymous and non-synonymous substitutions and the number of amino acid variants using the program DnaSP 5.10.01 (Librado & Rozas, 2009). To estimate population differentiation at *TLR2* we calculated pairwise F_{ST} (Weir & Cockerham, 1984) in Arlequin 3.5.1.3 (Excoffier & Lischer, 2010). Sequential Bonferroni corrections were applied to account for multiple testing. To test if patterns of haplotype frequencies deviated from neutral expectations we calculated Fay & Wu's *H* (Fay & Wu, 2000). The empirical distribution of the test statistics was generated using neutral coalescent simulations in DNAsp (Librado & Rozas, 2009), based on the observed number of segregating sites, 20 000 replicates and allowing for a recombination rate of 0.001. A *Mus musculus TLR2* sequence (NM_011905.3) was used as an outgroup.

We then reconstructed a *TLR2* haplotype genealogy in the program TCS 1.21 (Clement, Posada & Crandall, 2000) by taking synonymous and non-synonymous substitutions into account. We used the resulting genealogy to assign haplotypes to *TLR2* haplogroups (or clusters, defined as groups of haplotypes that are separated from other groups of haplotypes by at least seven mutations) following Tschirren et al. (2013). The three 'intermediate' haplotypes in the Italian populations were separated by more than seven mutations from the other, previously described clusters (c_1, c_2, c_3) (see Results) and were therefore treated as a forth cluster (c_4) in the analyses. We tested for deviations of *TLR2* cluster frequencies from Hardy–Weinberg equilibrium within populations in Arlequin 3.5.1.2. (Excoffier & Lischer, 2010). To determine to which of the previously described European bank vole

234 lineages (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006) the different study
235 populations belong, we compared the *cytb* sequences with 14 randomly
236 chosen *cytb* sequences from the seven lineages obtained from GenBank
237 (Basque: EF408066, EF408062; Spanish: EF408061, EF408060; Italian:
238 AJ639664, AJ639692; Western: DQ472319, DQ472333; Basal: DQ472327,
239 DQ472344; Eastern: DQ472347, DQ472346; Carpathian: DQ472253,
240 DQ472339). We excluded the eighth lineage (Ural) due to its large genetic
241 distance to all other lineages and because preliminary analyses showed that
242 we had no representative of this lineage in our sample. A neighbour-joining
243 tree (Maximum Composite Likelihood, 1000 bootstrap iterations) with
244 Northern red-backed vole, *Myodes rutilus* (Pallas, 1779; Rodentia, Cricetidae)
245 (JF714850) as an outgroup was created in MEGA 6 (Tamura *et al.*, 2013) and
246 used to assign the *cytb* sequences to one of the seven mitochondrial
247 lineages. To further confirm the robustness of the assignment of the study
248 populations to the different lineages, we reconstructed a *cytb* genealogy in the
249 program TCS 1.21 (Clement, Posada & Crandall, 2000) by taking
250 synonymous and non-synonymous substitutions into account. Finally, we
251 used a simulated annealing approach to identify groups that are maximally
252 differentiated from each other without taking geographical constrains into
253 account in the program SAMOVA 2.0 (Dupanloup *et al.,* 2002). This lineage
254 classification was then included in an analysis of molecular variance
255 (AMOVA) in Arlequin 3.5.1.3 (Excoffier & Lischer, 2010), which partitions the
256 total *TLR2* cluster variation into covariance fractions, resulting from the
257 variance among mitochondrial lineages, among populations within lineage,

and within populations. The number of permutations for this analysis was set to 1000.

Results

Mitochondrial lineages

The different bank vole populations could be assigned to one of five previously described mitochondrial lineages (Spanish, Italian, Western, Eastern, and Carpathian (Deffontaine *et al.*, 2005; Kotlík *et al.*, 2006) based on *cytb* sequencing (Figure 1; Figures S5, S6).

Groups that were maximally genetically differentiated from each other were identified as (SAMOVA; K = 5): Group 1 = ESP; Group 2 = ITA1, ITA2; Group 3 = NOR, SCO, SWE, LTU; Group 4 = RUS, UKR, DEN, FIN, POL, GER2; Group 5 = AUT, SUI, ENG, SLO, BEL, NED, GER1, CZE, which is identical to the lineage assignment of the populations based on the phylogenetic tree (Figure S5).

TLR2 haplotype diversity and differentiation

Genetic diversity at *TLR2* was high across bank vole populations with 86 haplotypes and 51 amino acid variants (Table 1). We observed a large number of private alleles within populations (N = 67, Table 1), even among populations in close proximity, as for example the two Italian populations ITA1 and ITA2 (Figure S7). Furthermore, we observed significant *TLR2* population

differentiation (F_{ST}) for most pairwise population comparisons (Table S11). An excess of high-frequency derived haplotypes was observed in most populations, which is indicative of positive selection (Table S8).

TLR2 clusters

TLR2 haplotypes from all European bank vole populations grouped in the same, distinct clusters (c_1, c_2, c_3, c_4; Figure 2, Tables S4, S9). Despite the high *TLR2* diversity, large number of private alleles and strong population differentiation (see above), haplotypes of the three main clusters (c_{1-3}) were found in the Western, Eastern, and Carpathian lineages, and two of the three main clusters (c_1, c_3) were observed in the Spanish lineage (Figure 1). Only haplotypes of one previously described cluster (c_1) were present in the Italian lineage. In addition, a fourth haplotype cluster, restricted to the Italian populations, was observed (c_4; Figures 1, 2).

An analysis of molecular variance (AMOVA) revealed that mitochondrial lineage explained 18% of variation in *TLR2* clusters, whereas population within lineage explained 14% of variation in *TLR2* clusters. Most variation in *TLR2* clusters was observed within populations (68%) (Table 2a). In comparison, lineage explained 55% of *cytb* variation (Table S10). When restricting the AMOVA analysis to the Western, Eastern and Carpathian lineages the proportion of variation in *TLR2* explained by lineage dropped to 0.8% (Table 2b), showing that the lineage effect in the overall analysis was due to differences in the *TLR2* composition of the Italian and Spanish clades. There was no evidence that *TLR2* cluster frequencies deviated from Hardy-

305 Weinberg equilibrium within populations (all *P* = non-significant after
306 Bonferroni correction).
307
308
309 Discussion

310 Molecular genetic analyses revealed a high *TLR2* diversity in 21 bank
311 vole populations across Europe. Although the haplotype composition differed
312 markedly among populations, and a large number of private alleles was
313 observed within populations, *TLR2* haplotypes from all populations grouped in
314 the same distinct clusters.
315 Haplotypes of the three main clusters were observed in the Western,
316 Eastern and Carpathian mitochondrial lineages, and it is possible that we
317 underestimated *TLR2* diversity in the two other lineages (i.e. the Italian
318 lineage where c_1 and c_4 occurred, and Spanish lineage where c_1 and c_3
319 occurred), because fewer samples were analysed for these groups. A
320 molecular analysis of variance confirmed that most *TLR2* cluster variation was
321 present within populations, rather than among populations or mitochondrial
322 lineages. Furthermore, the variance explained by lineage was due to the very
323 high frequency of c_3 in the Spanish lineage as well as the occurrence of a
324 fourth cluster that was restricted to the Italian lineage. When excluding the
325 Italian and Spanish clades, the variation explained by lineage dropped to
326 0.8% and was no longer significant. It shows that at least among the Eastern,
327 Western and Carpathian clades there is no association between *TLR2*
328 clusters and mitochondrial lineages.

329 Interestingly, the finding that *TLR2* clusters are shared among lineages
330 and populations across Europe is similar to patterns observed at the bank
331 vole MHC class II (*Dqa*-exon 2*)* where no genetic structure was observed at
332 the continental scale and mitochondrial lineage explained <2% of *Dqa*-exon 2
333 variation (Malé *et al.*, 2012). One interesting hypothesis is that parasites might
334 impose similar selective pressures on the two immune genes, thereby
335 maintaining genetic variation in time and space.
336 Our results suggest that the distinct *TLR2* clusters may precede the
337 split of the bank vole lineages, which is estimated to have occurred during the
338 Upper Pleistocene between 0.19 and 0.56 Mya (Deffontaine *et al.*, 2009), and
339 that clusters have been maintained within lineages since. Alternatively, *TLR2*
340 clusters may have arisen several times independently by convergent evolution
341 after the last glaciation. This, however, appears unlikely given the large
342 number of populations in which all three clusters were observed. Importantly,
343 both scenarios suggest that the co-occurrence of the three main *TLR2*
344 clusters may be favoured by natural selection. Haplotypes within clusters, on
345 the other hand, do not appear to be maintained across populations and
346 lineages, as suggested by the strong population differentiation and the large
347 number of private alleles within populations. This within-cluster differentiation
348 among population is likely the result of population isolation and drift, although
349 we can currently not exclude other evolutionary forces.
350 Only three haplotypes in the southernmost populations ITA1 and ITA2
351 did not group with one of the three previously described *TLR2* clusters. It is
352 possible that these haplotypes represent a forth cluster that was lost in other

353 populations during glaciation, but persisted on the Italian peninsula, which
354 remained largely free of ice (Hewitt 1999). Alternatively, the occurrence of
355 these 'intermediate' haplotypes could be an indication for relaxed selection in
356 southern Europe. Indeed, a strong association between *TLR2* genotype and
357 *Borrelia afzelii* resistance has recently been observed in bank voles (Tschirren
358 *et al.*, 2013), and whereas *Borrelia* sp. are common tick-transmitted
359 pathogens in most parts of Europe (Lindgren & Jaenson, 2006), they are
360 absent in central Italy where populations ITA1 and ITA2 are situated
361 (Cimmino *et al.*, 1992). Reduced parasite-mediated selection and / or
362 bottlenecks during post-glacial recolonization could also explain the low *TLR2*
363 diversity observed in populations at the northern periphery of the species
364 range (i.e. Norway and British Isles).
365 Interestingly, in Spain the *Borrelia* genospecies 'R57' is highly
366 prevalent (but absent in other parts of Europe) as is the otherwise rare *TLR2*
367 cluster c_3 (Gil *et al.*, 2005; Barandika *et al.*, 2007). Testing for interactive
368 effects between different *Borrelia* species and *TLR2* clusters will thus be a
369 fruitful next step to better understand why the distinct *TLR2* clusters are
370 maintained within bank vole lineages and populations, and why some *TLR2*
371 clusters are common in some populations, but rare in others. Also, TLR2 is
372 involved in the recognition of other pathogens (Lien et al., 1999; Medzhitov,
373 2001), which constitute multiple selective forces on TLR2 and may contribute
374 to the observed patterns of *TLR2* diversity and differentiation across Europe.
375 Besides interactions between host genotype and parasite strains or
376 species (i.e. host G x parasite G interactions), other processes could

contribute to the maintenance of *TLR2* clusters within populations. For example, overdominance appears to play a role in the maintenance of diversity at the major histocompatibility complex (MHC) (Doherty & Zinkernagel, 1975; Apanius *et al.*, 1997; Oliver, Telfer & Piertney, 2009), and has also been suggested to contribute to the maintenance of genetic variation at TIRAP, an adaptor in the TLR signalling pathway (Khor *et al.*, 2007). However, we found no evidence for an excess of *TLR2* heterozygotes in our study populations. Alternatively, negative-frequency dependent selection might underlie the maintenance of *TLR2* clusters within populations (Tanaka & Nei, 1989; Woolhouse *et al.*, 2002).

In conclusion, our study shows that *TLR2* haplotypes group in distinct clusters in bank vole populations across Europe. Because haplotypes of the main three *TLR2* clusters were found in at least three mitochondrial lineages, our results suggest that selection may maintain genetic supertypes at this innate immune receptor in time and space.

Acknowledgements

We thank Robert J. Baker from the Texas Tech University for providing Ukrainian bank vole samples, Irene Keller and three anonymous reviewers for comments on the manuscript, and the numerous helpers that assisted with bank vole trapping and DNA extraction in the different study populations. The study was financially supported by the University of Zurich Research Priority Program ‘Evolution in Action: From Genomes to Ecosystems’ and the Swiss National Science Foundation (PP00P3_157455)(to BT).

401 References

402 **Akira S, Uematsu S, Takeuchi O. 2006.** Pathogen recognition and innate
403 immunity. *Cell* **124:** 783-801.

404 **Andrés AM, Hubisz MJ, Indap A, Torgerson DG, Degenhardt JD, Boyko**
405 **AR, Gutenkunst RN, White TJ, Green ED, Bustamante CD, Clark**
406 **AG, Nielsen R. 2009.** Targets of balancing selection in the human
407 genome. *Molecular Biology and Evolution* **26:** 2755-2764.

408 **Apanius V, Penn D, Slev PR, Ruff LR, Potts WK. 1997.** The nature of
409 selection on the major histocompatibility complex. *Critical Reviews in*
410 *Immunology* **17:** 179-224.

411 **Barandika JF, Hurtado A, Garcia-Esteban C, Gil H, Escudero R, Barral M,**
412 **Jado I, Juste RA, Anda P, Garcia-Perez AL. 2007.** Tick-borne
413 zoonotic bacteria in wild and domestic small mammals in northern
414 Spain. *Applied and Environmental Microbiology* **73:** 6166-6171.

415 **Bernatchez L, Landry C. 2003.** MHC studies in nonmodel vertebrates: what
416 have we learned about natural selection in 15 years? *Journal of*
417 *Evolutionary Biology* **16:** 363-377.

418 **Bubb KL, Bovee D, Buckley D, Haugen E, Kibukawa M, Paddock M,**
419 **Palmieri A, Subramanian S, Zhou Y, Kaul R, Green P, Olson MV.**
420 **2006.** Scan of human genome reveals no new loci under ancient
421 balancing selection. *Genetics* **173:** 2165-2177.

422 **Cimmino MA, Fumarola D, Sambri V, Accardo S. 1992.** The epidemiology
423 of Lyme borreliosis in Italy. *Microbiologica* **15:** 419-424.

Clement M, Posada D, Crandall K. 2000. TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9:** 1657-1660.

Corbet GB, Harris S. 1991. *The Handbook of British Mammals*. Blackwell Scientific Publications: Oxford.

Deffontaine V, Ledevin R, Fontaine MC, Quere JP, Renaud S, Libois R, Michaux JR. 2009. A relict bank vole lineage highlights the biogeographic history of the Pyrenean region in Europe. *Molecular Ecology* **18:** 2489-2502.

Deffontaine V, Libois R, Kotlík P, Sommer R, Nieberding C, Paradis E, Searle JB, Michaux J. 2005. Beyond the Mediterranean peninsulas: evidence of central European glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys glareolus*). *Molecular Ecology* **14:** 1727-1739.

Doherty PC, Zinkernagel RM. 1975. Enhanced immunological surveillance in mice heterozygous at H-2 gene complex. *Nature* **256:** 50-52.

Drummond AJ, Ashton B, Cheung M, Heled J, Kearse M, Moir R, Stones-Havas S, Thierer T, Wilson A. 2009. Geneious v5.6.5.: Available from http://www.geneious.com.

Dupanloup, I., Schneider, S., Excoffier, L. 2002. A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology* **11**: 2571-2581.

Excoffier L, Lischer HEL. 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10:** 564-567.

448 **Fay JC, Wu CI. 2000** Hitchhiking under positive Darwinian selection. *Genetics*
449 **155**, 1405–1413.
450 **Ferguson W, Dvora S, Fikes RW, Stone AC, Boissinot S. 2012.** Long-term
451 balancing selection at the antiviral gene OAS1 in Central African
452 chimpanzees. *Molecular Biology and Evolution* **29:** 1093-1103.
453 **Fornůsková A, Bryja J, Vinkler M, Macholán M, Piálek J. 2014.** Contrasting
454 patterns of polymorphism and selection in bacterial-sensing toll-like
455 receptor 4 in two house mouse subspecies. *Ecology and Evolution* **4:**
456 2931-2944.
457 **Gil H, Barral M, Escudero R, Garcia-Perez AL, Anda P. 2005.** Identification
458 of a new Borrelia species among small mammals in areas of northern
459 Spain where Lyme disease is endemic. *Applied and Environmental*
460 *Microbiology* **71:** 1336-1345.
461 **Guan Y, Ranoa DRE, Jiang S, Mutha SK, Li XY, Baudry J, Tapping RI.**
462 **2010.** Human TLRs 10 and 1 share common mechanisms of innate
463 immune sensing but not signaling. *Journal of Immunology* **184:** 5094-
464 5103.
465 **Haldane JBS. 1949.** Disease and evolution. *La Ricerca Scientifica* **19:** 68-76.
466 **Hedrick PW. 1998.** Balancing selection and MHC. *Genetica* **104:** 207-214.
467 **Hedrick PW. 2002.** Pathogen resistance and genetic variation at MHC loci.
468 *Evolution* **56:** 1902-1908.
469 **Hewitt GM.** 1999. Post-glacial re-colonization of European biota. *Biological*
470 *Journal of the Linnean Society* **68:** 87-112.

Jaarola M, Tegelström H, Fredga K. 1999. Colonization history in Fennoscandian rodents. *Biological Journal of the Linnean Society* **68:** 113-127.

Jackson JA, Friberg IM, Bolch L, Lowe A, Ralli C, Harris PD, Behnke JM, Bradley JE. 2009. Immunomodulatory parasites and toll-like receptor-mediated tumour necrosis factor alpha responsiveness in wild mammals. *BMC Biology* **7:** doi:10.1186/1741-7007-1187-1116.

Jin MS, Kim SE, Heo JY, Lee ME, Kim HM, Paik SG, Lee HY, Lee JO. 2007. Crystal structure of the TLR1-TLR2 heterodimer induced by binding of a tri-acylated lipopeptide. *Cell* **130:** 1071-1082.

Kang JY, Nan X, Jin MS, Youn SJ, Ryu YH, Mah S, Han SH, Lee H, Paik SG, Lee JO. 2009. Recognition of lipopeptide patterns by Toll-like receptor 2-Toll-like receptor 6 heterodimer. *Immunity* **31:** 873-884.

Khor CC, Chapman SJ, Vannberg FO, Dunne A, Murphy C, Ling EY, Frodsham AJ, Walley AJ, Kyrieleis O, Khan A, Aucan C, Segal S, Moore CE, Knox K, Campbell SJ, Lienhardt C, Scott A, Aaby P, Sow OY, Grignani RT, Sillah J, Sirugo G, Peshu N, Williams TN, Maitland K, Davies RJO, Kwiatkowski DP, Day NP, Yala D, Crook DW, Marsh K, Berkley JA, O'Neill LAJ, Hill AVS. 2007. A Mal functional variant is associated with protection against invasive pneumococcal disease, bacteremia, malaria and tuberculosis. *Nature Genetics* **39:** 523-528.

Kotlík P, Deffontaine V, Mascheretti S, Zima J, Michaux JR, Searle JB. 2006. A northern glacial refugium for bank voles (*Clethrionomys*

glareolus). *Proceedings of the National Academy of Sciences of the USA* **103:** 14860-14864.

Laird PW, Zijderveld A, Linders K, Rudnicki MA, Jaenisch R, Berns A. 1991. Simplified mammalian DNA isolation procedure. *Nucleid Acids Research* **19:** 4293.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25:** 1451-1452.

Lien E, Sellati TJ, Yoshimura A, Flo TH, Rawadi G, Finberg RW, Carroll JD, Espevik T, Ingalls RR, Radolf JD, Golenbock DT. 1999. Toll-like receptor 2 functions as a pattern recognition receptor for diverse bacterial products. *Journal of Biological Chemistry* **274**: 33419-33425.

Lindgren E, Jaenson TGT. 2006. Lyme borreliosis in Europe: influences of climate and climate change, epidemiology, ecology and adaptation measures. In: Europe WRO, ed. Copenhagen: World Health Organisation. 1-35.

Malé PJG, Martin JF, Galan M, Deffontaine V, Bryja J, Cosson JF, Michaux J, Charbonnel N. 2012. Discongruence of *Mhc* and cytochrome *b* phylogeographical patterns in *Myodes glareolus* (Rodentia: Cricetidae). *Biological Journal of the Linnean Society* **105:** 881-899.

Medzhitov R. 2001. Toll-like receptors and innate immunity. *Nature Reviews Immunology* **1:** 135-145.

Morger J, Bajnok J, Boyce K, Craig PS, Rogan MT, Lun ZR, Hide G, Tschirren B. 2014. Naturally occurring Toll-like receptor 11 (*TLR11*)

519 and Toll-like receptor 12 (*TLR12*) polymorphisms are not associated
520 with *Toxoplasma gondii* infection in wild wood mice. *Infection Genetics*
521 *and Evolution* **26**: 180-184.
522 **Netea MG, Wijmenga C, O'Neill LAJ. 2012.** Genetic variation in Toll-like
523 receptors and disease susceptibility. *Nature Immunology* **13:** 535-542.
524 **Oliver MK, Telfer S, Piertney SB. 2009.** Major histocompatibility complex
525 (MHC) heterozygote superiority to natural multi-parasite infections in
526 the water vole (*Arvicola terrestris*). *Proceedings of the Royal Society B*
527 **276:** 1119-1128.
528 **Piertney SB, Oliver MK. 2006.** The evolutionary ecology of the major
529 histocompatibility complex. *Heredity* **96:** 7-21.
530 **Stephens M, Smith NJ, Donnelly P. 2001.** A new statistical method for
531 haplotype reconstruction from population data. *American Journal of*
532 *Human Genetics* **68:** 978-989.
533 **Takahata N, Nei M. 1990.** Allelic genealogy under overdominant and
534 frequency-dependent selection and polymorphism of Major
535 histocompatibility complex loci *Genetics* **124:** 967-978.
536 **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013.** MEGA6:
537 Molecular Evolutionary Genetics Analysis Version 6.0. *Molecular*
538 *Biology and Evolution* **30:** 2725-2729.
539 **Tanaka T, Nei M. 1989.** Positive Darwinian selection observed at the
540 variable-region genes of immunoglobulins *Molecular Biology and*
541 *Evolution* **6:** 447-459.

Texereau J, Chiche JD, Taylor W, Choukroun G, Comba B, Mira JP. 2005. The importance of Toll-like receptor 2 polymorphisms in severe infections. *Clinical Infectious Diseases* **41:** S408-S415.

Tschirren B, Andersson A, Scherman K, Westerdahl H, Mittl P, Råberg L. 2013. Polymorphisms at the innate immune receptor *TLR2* are associated with *Borrelia* infection in a wild rodent population *Proceedings of the Royal Society B* **280:** 10.1098/rspb.2013.0364.

Tschirren B, Andersson M, Scherman K, Westerdahl H, Råberg L. 2012. Contrasting patterns of diversity and population differentiation at the innate immunity gene Toll-like receptor 2 (*TLR2*) across populations of two sympatric rodent species. *Evolution* **66:** 720-731.

Tschirren B, Råberg L, Westerdahl H. 2011. Signatures of selection acting on the innate immunity gene Toll-like receptor 2 (*TLR2*) during the evolutionary history of rodents. *Journal of Evolutionary Biology* **24:** 1232-1240.

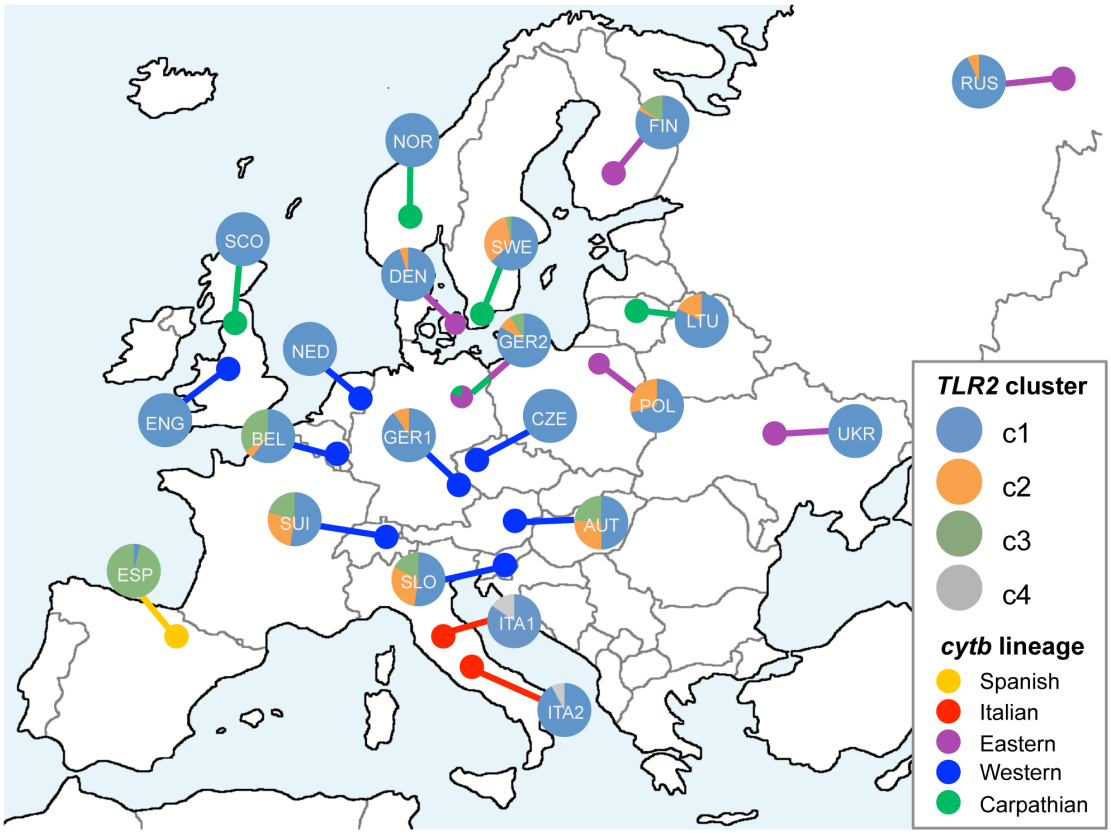
Turner AK, Begon M, Jackson JA, Bradley JE, Paterson S. 2011. Genetic diversity in cytokines associated with immune variation and resistance to multiple pathogens in a natural rodent population. *PLoS Genetics* **7:** e1002343. doi:1002310.1001371/journal.pgen.1002343.

Weir BS, Cockerham CC. 1984. Estimating F-statistic for the analysis of population-structure. *Evolution* **38:** 1358-1370.

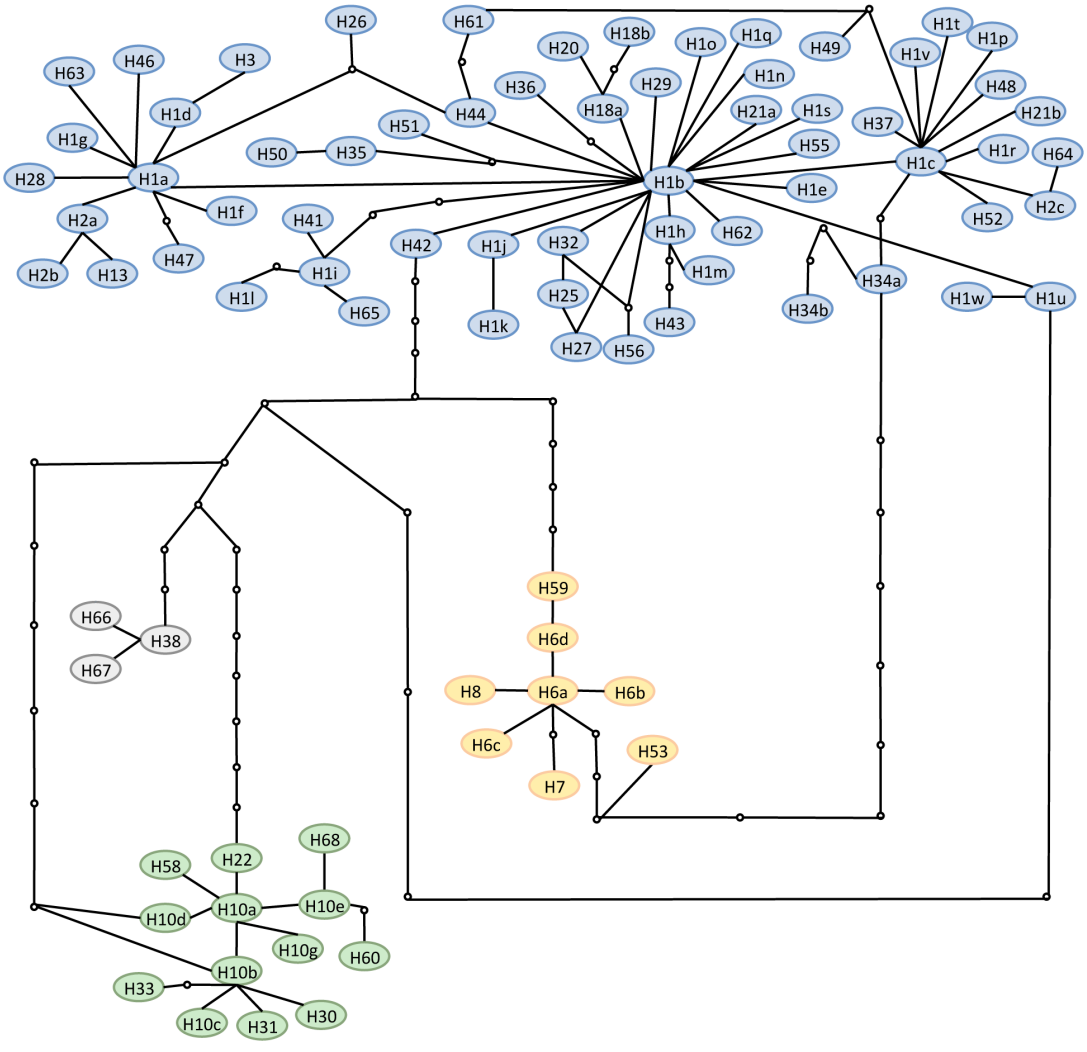
Woolhouse MEJ, Webster JP, Domingo E, Charlesworth B, Levin BR. 2002. Biological and biomedical implications of the co-evolution of pathogens and their hosts. *Nature Genetics* **32:** 569-577.

Figure Legends

Figure 1. *TLR2* cluster frequencies and mitochondrial lineage of the sampled bank vole populations. Population IDs are given within the corresponding *TLR2* cluster pie. N = 385 animals; N = 20 – 42 alleles per population.



572 Figure 2. *TLR2* haplotype network. Haplotypes that differ at the amino acid
573 level are labelled with different numbers (e.g. H1, H2, and H3). Haplotypes
574 that do not differ at the amino acid level are labelled with different letters (e.g.
575 H1a, H1b, and H1c). Haplotypes are coloured according to clusters: c_1 (blue),
576 c_2 (orange), c_3 (green), c_4 (grey). Each line indicates a change at one
577 nucleotide.



578